

Using GIS to Study the Health Impact of Air Emissions

Andrew L Dent (1),* David A Fowler (2), Brian M Kaplan (3), Gregory M Zarus (4)
(1) GIS Analyst, Electronic Data Systems, Plano, TX; (2) Toxicologist, Agency for Toxic Substances and Disease Registry, Exposure Investigations and Consultation Branch, Atlanta, GA; (3) Environmental Health Scientist, Agency for Toxic Substances and Disease Registry, Federal Facilities Assessment Branch, Atlanta, GA; (4) Atmospheric Scientist, Agency for Toxic Substances and Disease Registry, Exposure Investigations and Consultation Branch, Atlanta, GA

Disclaimer

The use of company or product names is for identification only and does not constitute endorsement by the Agency for Toxic Substances and Disease Registry or the US Department of Health and Human Services.

Abstract

Geographic information system (GIS) technology is fast-developing with an ever-increasing number of applications. Air dispersion modeling is a well-established discipline that can produce results in a spatial context. The marriage of these two applications is optimal because it leverages the predictive capacity of modeling with the data management, analysis, and display capabilities of GIS. In the public health arena, exposure estimation techniques are invaluable. The utilization of air emission data, such as the US Environmental Protection Agency's Toxics Release Inventory data, and air dispersion modeling with GIS enable public health professionals to identify and define a potentially exposed population, estimate the health risk burden of that population, and determine correlations between point-based health outcome results and estimated health risk.

Keywords: air pollution, emissions, toxics release inventory, public health, Air Force

Introduction

The federal Agency for Toxic Substances and Disease Registry (ATSDR) is often charged with investigating past exposures to determine their associations with health outcomes observed within specific communities. The task often requires the use of information specific to a given locality, namely: toxic substance release data, topographical data, land use data, meteorological data, mathematical modeling data, population density data, demographic data, and health outcome data. Proper evaluation and correlation of these data are essential to reveal possible associations between observed health effects and chemical exposures. This paper details a technique by which air dispersion model results are integrated with other spatial datasets on a geographic information system (GIS) platform. Once the component datasets are gathered, they can be

* Andrew L Dent, Electronic Data Systems, 5400 Legacy Dr., Plano, TX 75024 USA; (p) 404-639-6099; E-mail: aed5@cdc.gov

analyzed using standard GIS functions to provide support for interpretation of site data and evaluation of exposure hypotheses.

Public Health Context

The analytical technique explained in this paper will be showcased in the context of a public health concern brought on by environmental contamination. The public health concern existed at a US Air Force base charged with the management and maintenance of aircraft engines, weapons systems, support equipment, and aerospace fuels. The base hosted, maintained, and repaired various jet aircraft. Specific activities that were potential sources of off-site air contamination included painting, chrome plating, fueling, and fuel storage. Members of the community neighboring the base expressed concern about fuel vapor and other odors, and questioned the relationship between the odors and the occurrence of health effects such as nausea, headaches, difficulty breathing, and cancer.

Integration Technique

The technique integrates conventional air dispersion modeling with GIS technology. The marriage of these two technologies is appropriate because it takes advantage of each component's strengths—the high-powered, predictive capacity of computer models and the editing, data handling, and interpolation capabilities of GIS (1). This paper first explains the basics of air dispersion modeling, including its development, uses, and advantages. Next, it outlines the rise of GIS technology and investigates a variety of GIS applications. Finally, this paper concludes with a discussion of the integration of air dispersion modeling results with other spatial data on a GIS platform.

Introduction to Air Dispersion Modeling

Air dispersion modeling is a predictive method used to estimate the concentrations of pollutants in the atmosphere resulting from point or nonpoint atmospheric emissions. It takes into account various factors that can affect a substance's concentration in a plume as it migrates through the air. These influential factors include gravity, meteorological conditions, and chemical reactions. Among a variety of other purposes, air dispersion models have been used to:

- Design stacks to minimize the nuisance from pollutants at ground level
- Calculate when odors might be expected
- Determine the needed removal efficiency of air pollution control equipment
- Plan for emergency response to accidental releases
- Determine the acceptable levels of operation of air pollution generating facilities (2)

The regulatory role of air dispersion models has gained great significance because judicial rulings in 1977, 1978, and 1980 have upheld the Clean Air Act Amendment of 1977, which states that a modeled air pollution violation is just as valid as a sampled violation (2).

The Creation of Air Models

Air dispersion models are semiempirical—they are in part derived from basic

principles and in part derived from measured data. Specifically, an air dispersion model is developed by measuring both emissions and meteorological conditions and monitoring contaminant concentrations at various points downwind of the source. These data are then correlated to find a mathematical model equation that provides the best fit. The resulting model is validated through rigorous testing at various combinations of known emissions and meteorological conditions. Although the mathematics of the model greatly affect its successful prediction of a substance's concentration, the quality of the input data also has a great impact on the results (2).

Advantages of Air Dispersion Modeling

The use of air modeling is advantageous for a variety of reasons, including the following (3):

- Models can be used to estimate a substance's concentration 24 hours a day for any time period for which both emissions and meteorological data exist.
- Models can account for deposition from particulate matter settling to the ground, as well as depletion resulting from the substance's reaction with sunlight and other materials.
- Models can be used to estimate the level of various substances existing in the ambient air as a result of emissions from a single source or multiple sources.
- Models can average short-term fluctuations in emissions and meteorological conditions, resulting in a long-term average.
- Models can estimate a substance's concentration at an unlimited number of locations.

Conversely, conventional air sampling can be limiting for a variety of reasons, including the following:

- Sampling measures substances arising from many and varied sources in the area; it cannot determine the effects of a single facility.
- Sampling results are based on conditions at the time of the sampling event. These conditions could be an extreme and not represent average conditions (3).
- Sampling efforts can be very expensive.

Geographic Information System Technology

GIS technology provides an excellent platform upon which different types of spatially referenced data can be united for analysis and display purposes. Prior to the advent of GIS technology, many operations that involved the concerted utilization of datasets derived from different sources and in different formats were carried out using a "pushpin" approach in which hard-copy maps were generated and overlaid upon one another. The approach was both costly and time-consuming and often yielded substandard results. This type of spatial analysis harkens back to a map by French military leader and cartographer Louis Alexandre Berthier (1753–1815) that was composed of hinged overlays showing troop movements during the 1781 Siege of Yorktown (4). In the 1960s, Dr Roger Tomlinson of the Canada Geographic Information System (CGIS) developed the first computer-based GIS (5). CGIS was designed to store, manage,

analyze, and manipulate spatial data in an effort to assess the productivity of Canadian farmland.

The arrival of high-powered computers in the 1980s facilitated the proliferation of GIS applications in a variety of different disciplines. It has been estimated that over 80% of the world's data have a spatial component. These spatial components could include an address in a database, coordinates in sampling data, or a zip code in sales data. GIS has been used to manage county tax records, route delivery and emergency vehicles, perform site selection based on many parameters, manage utility networks, and develop strategies to address problems such as crime, urban sprawl, and environmental degradation.

Elements of a GIS

The process of developing a GIS includes data acquisition and preprocessing; data management, manipulation and analysis; and product generation (4). Data acquisition is often the most expensive and time-consuming element in utilizing GIS technology. Issues involved in this phase of development include data accuracy, scale, and metadata. Metadata is a written detailed description of the data similar to engineering specifications. Preprocessing, or the machinations necessary to convert data to a digital format and integrate it with other spatial data, often involves elements such as digitizing and quality assurance and control procedures. As with air dispersion modeling efforts, the successful use of GIS technology depends much upon the datasets used and the methods used to automate those datasets. Thus, data management, including storage and documentation, is key to a GIS project. During the manipulation and analysis phase, the data are used to get results and make decisions. It is important to remember that GIS is a tool—a flexible, easy to use tool—but, still, only a tool. The application or effective use of GIS is key to the success of a GIS enterprise. Gigabytes of spatial data are of no value to any organization if that information is not used. Finally, product generation involves the transmittal of results produced by GIS to the people who need them. These results may be delivered as a conventional paper map or digitally via diskettes, intranet, or Internet. Whereas the paper map is static, a digital map can be dynamic because it can contain a greater amount of information. This allows the map reader to define the map message by selecting different map scales or different data to display.

Integration of Modeling Results on a GIS Platform

Typically, GIS and air dispersion modeling software are separate packages often written in different languages. Therefore, the question arises as to how the air dispersion modeling results can be most effectively integrated with other spatial data on a GIS platform. Models for the integration of modeling results and GIS data include full integration, loose coupling, and tight coupling (Figure 1) (1). Full integration means that the calculations performed by the model are encoded in a high-level language packaged with the GIS software, such as ARC/INFO's Arc Macro Language (AML) or ArcView's Avenue (Environmental Systems Research Institute, Redlands, CA). Loose coupling integration consists of uniting the systems at predefined end points. For example, the data would first be processed using air modeling software, then the results would be used in a GIS package. Tight coupling integration involves the development of a user

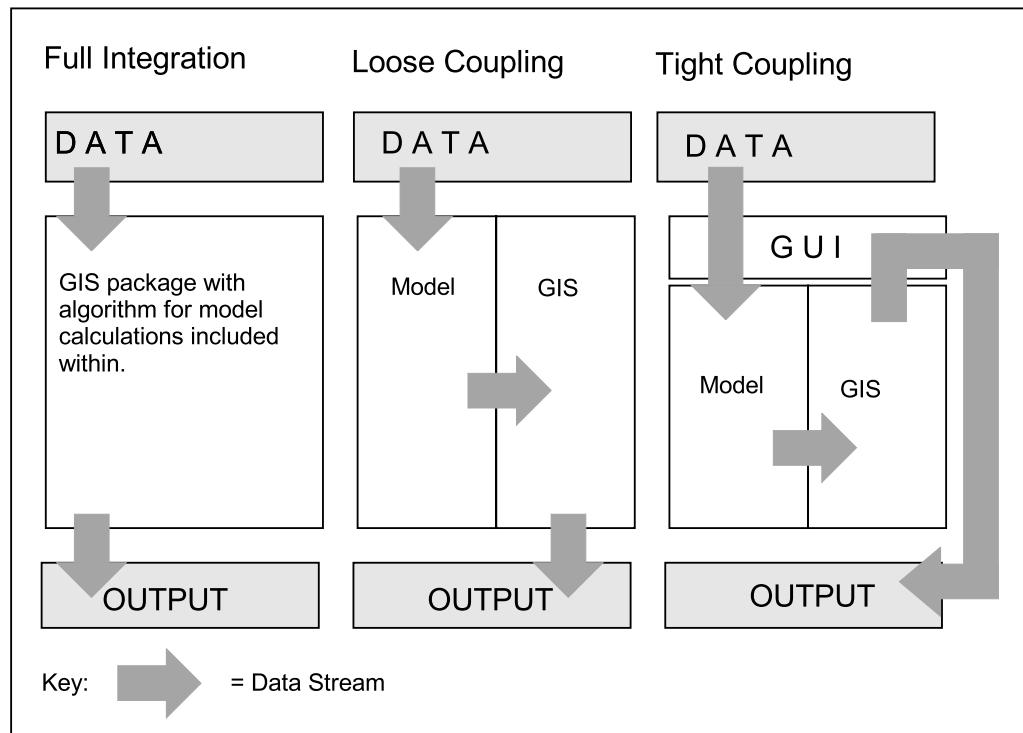


Figure 1 Integration strategies (1).

interface that allows users to access both the model software and the GIS package using one graphical user interface (GUI). An application built on the tight coupling concept gives the illusion that one software package is being used when, in reality, the GIS package and the modeling program are being used separately in concert with each other.

The work discussed in this paper uses loose coupling because loose coupling provides a faster implementation time than tight coupling and with the same results. If, in the future, this process needs to be used on a repetitive basis, it may be worth the time to develop a tight coupling integration application. In such a case, the investment in development time and effort would be offset by the speed and efficiency with which the task could then be carried out.

GIS Datasets

Use of air dispersion model results with other GIS spatial datasets on a GIS platform is at the heart of the work discussed here. This section discusses the spatial datasets that were compiled and integrated with the air dispersion model results. The conditions at the site and the goal of ATSDR in its investigation have warranted the use of various datasets. These include aerial photography, 1995 TIGER/Line files, and US Census demographic data.

Aerial Photography

Aerial photography can be gathered from a variety of data sources. Although image

processing techniques provide many ways to view and manipulate aerial photography, the authors used it here primarily for orientation purposes.

TIGER/Line Data

Topologically Integrated Geographically Encoded and Referencing (TIGER/Line) files (6) are spatial datasets compiled by the US Census Bureau that include base map features for all areas of the United States. These features include roads, rivers, water bodies, political boundaries, and cultural features (schools, parks, and hospitals). GIS data consisting of on-base features in large scale were obtained from the engineering division of the Air Force base; however, the aerial extent of the analysis to be performed required more extensive use of the TIGER/Line files.

US Census Demographic Data

ATSDR's work depends heavily on demographic information when analyzing environmental contamination and the potentially related health effects. Specifically, obtaining the raw numbers of people in an area and the numbers of people who are in high-risk groups (children, the elderly, and women of reproductive age) is critical. The US Census provides this type of data along with an abundance of other information on socioeconomic variables. For analysis at the Air Force base, block-level spatial data and demographic data were compiled. This information was dynamically integrated with the model-based results data in the GIS analysis phase of the project.

Integration Technique

ATSDR used the most recent version of the US Environmental Protection Agency's (EPA's) Industrial Source Complex Short Term air dispersion model, version 3 (ISCST3), to estimate emissions from the Air Force base. EPA developed the ISCST3 model primarily for determining if air emissions sources meet state and federal air quality standards (7). The ISCST3 model allows the refinement of results through the use of additional parameters such as building height, source temperature, particle size, and decay rate. When modeling the emissions from the Air Force base, ATSDR did not specify the values for these additional parameters. The default values were deemed appropriate because the potential improvement of the results arrived at through the setting of these parameters was considered to be negligible when compared to the uncertainty of the input parameters.

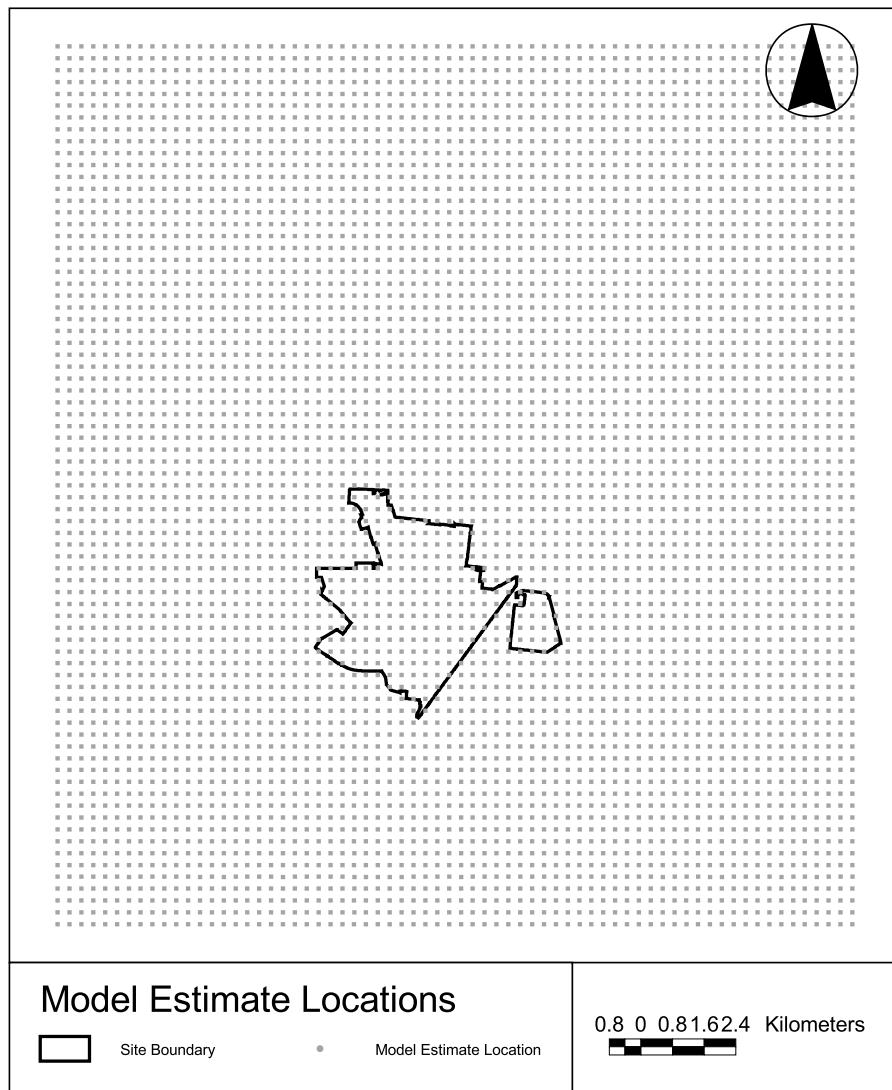
The model was run on 7,016 inputs (sources and multiple chemicals from each source) from the Air Force base emission inventory for a variety of carcinogenic compounds. This inventory, which is similar to EPA's Toxics Release Inventory (TRI), contains an emission rate in grams per second for each emitted air pollutant from each source. The model estimated concentration values for a uniform grid of 5,100 points spread evenly across the 446-square-kilometer study area at 300-meter intervals. Five years of meteorological data were averaged to form the meteorological component. The results included the coordinates for each modeled point, the average compound concentration estimate at each point, and the number of hours for which the compound concentration values were estimated.

A GIS point file for each of the modeled contaminants was created. One point in the GIS file exists for each modeled grid point in the original air model output. Thus, the

spatially referenced modeled output dataset can now be integrated with other spatial datasets (Figure 2).

Map Projections

The air dispersion model produces contaminant values at grid intersections of a Cartesian coordinate system. This coordinate system is a two dimensional representation of the earth's surface. The coordinate system is derived from a map projection that is a mathematical method for representing the features of the spherical surface of the earth on a planar map. GIS datasets can be stored using a variety of different map



Source: ATSDR, 1998.

Figure 2 Locations of points at which the model generated a concentration estimate (3).

projections. For datasets to be integrated, however, it is imperative that they are stored in the same projection. Therefore, the integration of air dispersion modeling data involves the reprojection of data from the original map projection to the projection of the remaining datasets. In this case, the GIS point file was projected from the Texas State Plane South Central coordinate system (based on the Lambert conformal conic projection) to the Plate Carrée projection. After the importation and reprojection have been accomplished, the modeled datasets can be analyzed in concert with existing data.

Applicable GIS Processes

Once data have been integrated into a single digital map on a GIS platform, several spatial processes are available for analyzing the information. For instance, contours can be generated from values at point locations by a variety of interpolation techniques. Kriging is considered the optimal method of spatial linear interpolation. Kriging is a logarithmic method in which the mean is estimated from the best linear-weighted moving average (8). For the work presented in this paper, the authors used kriging to generate contours from the air dispersion modeling data because logarithmic interpolation is more consistent with the natural distribution of contamination.

Contours developed with kriging can be used to generate polygons that, in turn, can be used in additional GIS analysis, such as the area-proportion technique. The area-proportion technique is an excellent example of the use of GIS to analyze disparate types of data to get results. It is essentially a “cookie-cutter” operation that estimates values within a polygon based on values of polygons in another data layer. For instance, the number of people living within a contour generated by kriging could be estimated based on the number of people living in census blocks in the same area. For cases in which the contour polygon crosses the census blocks, a simple proportion of the area of the census blocks lying within the target polygon is used to compute population numbers (Figure 3). The area-proportion technique assumes an even distribution of population (or whatever is being estimated) and, therefore, might result in a certain amount of measurable error. For example, the area-proportion technique assumes that the population within a census block is evenly distributed throughout the block. Thus, if a contour polygon encompasses 50% of a block it is assumed that 50% of the population of that block lives within the contour polygon. However, because population is rarely evenly distributed across an area, the area-proportion technique results in some amount of error. For example, if all of the population living in a census block that the contour polygon boundary intersects actually live *outside* the contour polygon, then the

16	32	8
16	48	16

The value associated with the grey square can be computed by:

$$(32/4) + (8/4) + (48/4) + (16/4) = 26$$

Figure 3 Area-proportion technique illustrated.

area-proportion estimate is high. Conversely, if all of the population living in a census block intersected by the contour polygon boundary actually live *inside* the buffer polygon, the area-proportion estimate is low. The size of the error generated by the area-proportion technique depends on:

- The number of reference polygons (e.g., census blocks) used to compute the estimate
- The size of the reference polygons

The application of the area-proportion technique is an excellent way to begin exploration of disparate datasets using GIS.

Public Health Applications

After all datasets have been melded in a GIS format, many avenues of data exploration are open to public health professionals. As stated earlier, ATSDR's goal is to evaluate the potential detrimental effects of air releases on the neighboring populations. This goal provides the impetus for integrating air dispersion modeling data with existing spatial datasets.

Calculation of Cancer Health Risk

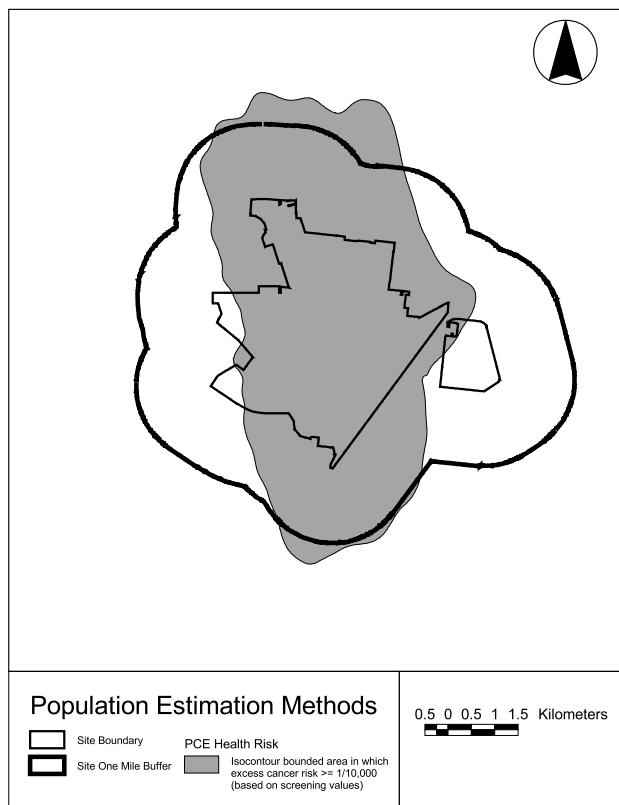
The additional cancer health risk to a population can be estimated by multiplying the EPA's cancer inhalation slope factor (unit risk) by the average annual concentration predicted by the model. This estimate is considered a screen because it is a worst-case conservative estimate and indicates which segments of the population of interest can be eliminated from further analysis and which segments require more refined analysis. Segments of the population requiring further analysis can be more accurately identified using this integrated approach, not only as to physical location, but also as to demographic composition. This process would allow a more site-specific approach to account for the prevalence of more sensitive subpopulations. This refined information can also be used to identify residents for information mailings and notifications, solicitation of information, and clinical intervention or medical monitoring.

Refinement of the Exposed Population

Without the use of modeling (or sampling), public health professionals are tied to obsolete methods for estimating the exposed population. For instance, some organizations might have specified that a population living within a uniform distance of a site boundary is the "exposed population." Population estimates derived in this manner are often substantially different from the actual exposed population. Modeling allows the refinement of size, location, and demographic composition of an exposed population. In the example shown in Figure 4, air dispersion modeling has resulted in a smaller exposed population than the estimate arrived at using the uniform distance technique (Table 1). Furthermore, the population derived from air modeling is distributed across a slightly different area.

Correlation with Point-Based Health Outcome Data

Often health professionals can obtain spatially referenced health outcome data for specific populations. Such a dataset would include the coordinates of a residence or



Source: ATSDR, 1998.

Figure 4 Exposed population estimation methods.

Table 1 Estimates of Population Exposed to Perchloroethene (Unnamed Air Force Base)

Area of Estimation	Total Population ^a
Population within 1 mile of site boundary	50,861
Population within 1/10,000 risk contour	26,033

^a Computed by area-proportion technique.

Source: (3,6)

workplace of an individual and the health outcome exhibited by that individual. Obtaining such data is often made difficult because of confidentiality issues; however, when they can be obtained, they can be successfully integrated with GIS data (such as air dispersion modeling results).

After the coordinates are imported into a GIS package, the health outcome points can be correlated with the contours of elevated health risk. This type of analysis can indicate if there is a potential association between a modeled exposure level and a reported health outcome. However, many other issues related to the makeup of the population must be considered. These critical elements include, but are not limited to:

- Length of residence

- Potential occupational exposure
- Potential residential exposure
- Genetics
- Socioeconomic status
- Lifestyle (e.g., smoking and nutrition)

Each of these elements can have an effect on the interpretation of health outcome data. For example, length of residence is considered because cancer development usually involves a latency period of many years. Therefore, if a person in the community of interest developed a cancer, but had only lived in the community a short time, it would be unlikely that the cancer development occurred as a result of an exposure in the community of interest.

Correlation Methods

At this time, the health outcome data based on the residence location of reported cases have not yet been compiled for the project discussed in this paper. Nonetheless, the analysis will generally proceed as follows.

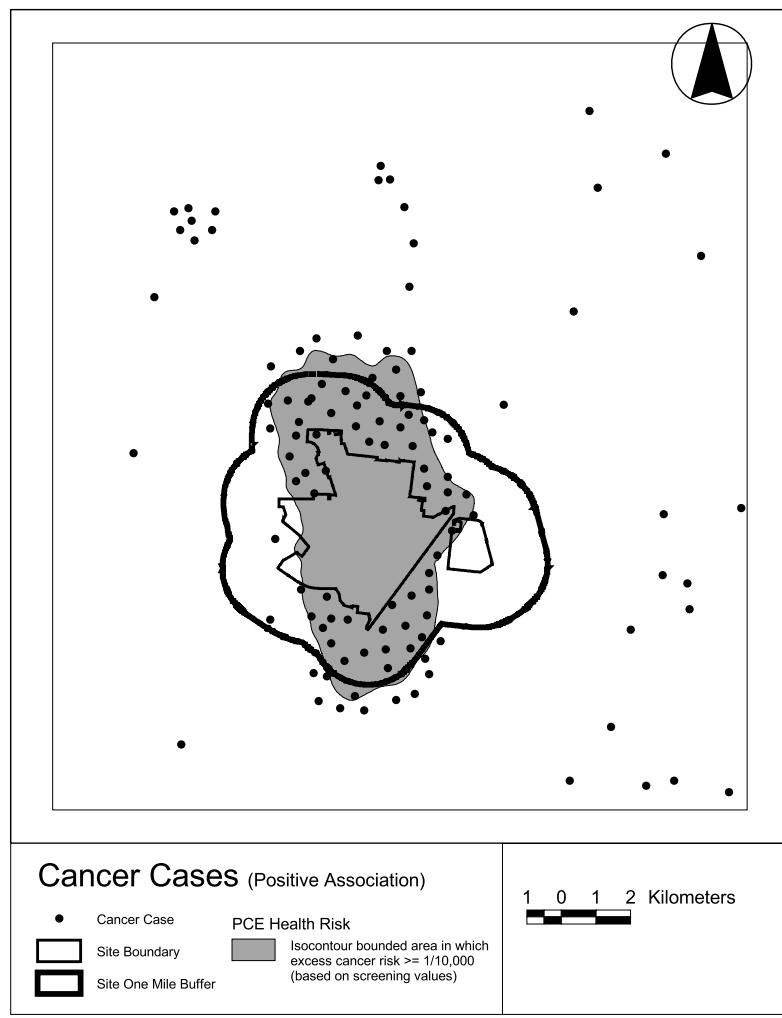
In determining the correlation of health outcome points with a cancer risk contour, five properties of the distribution of point features are pertinent. These are:

- Frequency: Number of occurrences
- Density: Number of occurrences per unit area
- Geometric center: Means of x-y coordinates
- Spatial dispersion: Standard deviation of the means of x-y coordinates
- Spatial arrangement: "Pattern" of the points, can be clustered, random, or scattered (8)

Based on the distribution of cancer case points in the vicinity of the study area, a judgement can be made on the potential association of the cases with the cancer risk contour. This judgement is just one factor in a weight-of-evidence approach that evaluates all relevant data, rather than a purely quantitative approach in which assumptions and uncertainties are often not represented (9).

The properties of point distribution can be utilized to evaluate the distribution of cases in many and varying ways. For instance, density can be used to establish association when the point density inside a critical contour is higher than the point density outside the contour. However, this measure should not be taken by itself. It must be normalized by population to yield a valid assessment. Furthermore, the spatial arrangement must be evaluated to completely assess the exposed population. For example, the clustering of cases at one or more points within the critical contour might be related to the location of a retirement or assisted-living home, not the exposure of a population to carcinogens. Finally, a comparison of the spatial arrangement of the cases inside and outside the critical contour can be enlightening. A marked difference (Figure 5), in which a scattered pattern exists inside the critical contour and a random or clustered pattern exists outside, could be an indication of a potential association of cases with chemical or substance exposure.

The distribution of cancer point cases can also be used to eliminate any association and accompanying public fears. For example, if the majority of cases lie outside the critical contour, it would be safe to infer that the polluter might not be emitting quantities



Source: ATSDR, 1998.

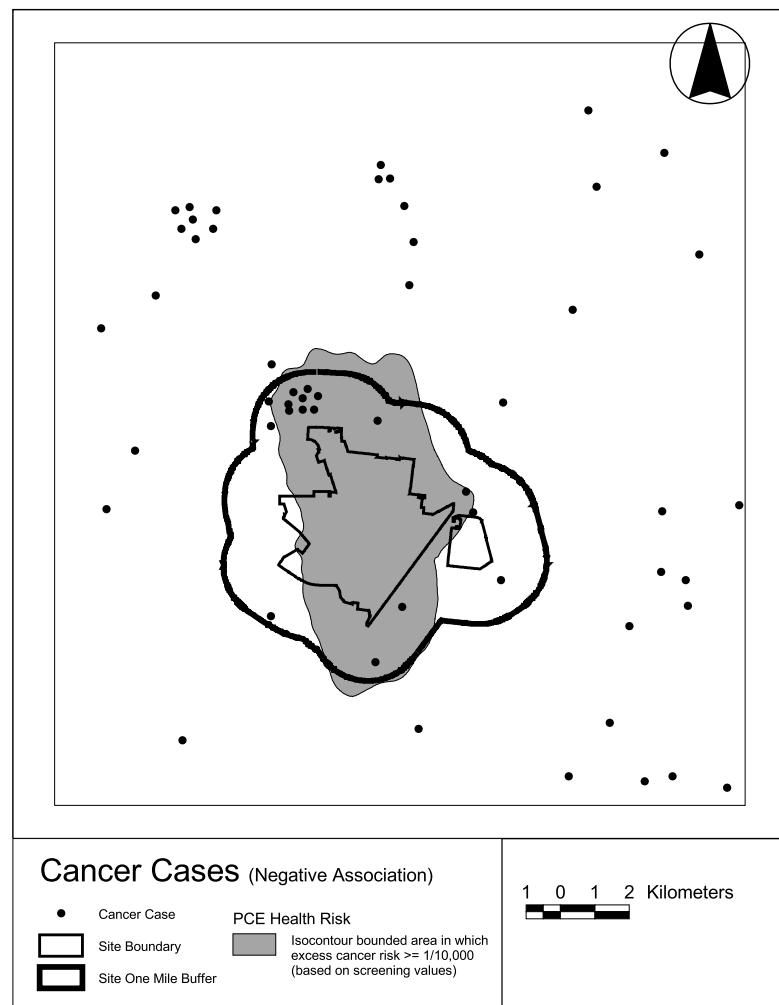
Figure 5 Positive association scenario.

of pollutants in a sufficient amount to be dangerous. In addition, as illustrated in Figure 6, a pattern that remains the same inside and outside the critical contour indicates that the cases may not be associated with exposure. Finally, a uniform density (normalized by population) can also be a reason to conclude that emissions may not be a factor in the health outcome.

The key to evaluation of case locations in the context of a critical risk contour is to consider the totality of many elements including, but not limited to, spatial distribution of points, population density, and population characteristics.

Conclusion

GIS is a fast-developing technology with an ever-increasing number of applications. Air



Source: ATSDR, 1998.

Figure 6 Negative association scenario.

dispersion modeling is a well-established discipline that can produce results in a spatial context. The marriage of these two applications is optimal because it leverages the predictive capacity of modeling and the data management, analysis, and display capabilities of GIS.

In the public health arena, exposure estimation techniques are invaluable. The use of air emission data, such as TRI data, and air dispersion modeling with GIS, enable public health professionals to identify a potentially exposed population, estimate the health risk burden of that population, and attempt to correlate point-based health outcome results with estimated health risk. As demonstrated in this paper, the GIS platform provides a means by which air dispersion modeling results can be effectively applied to public health studies.

References

1. Folgert D. 1998. An investigation of dynamic simulation: The integration of ARC/INFO with air dispersion modeling to facilitate data handling and visualization. In: *ESRI Eighteenth Annual User Conference proceedings*. Environmental Systems Research Institute, Inc., Redlands, CA.
2. Schulze RH, Turner DB. 1996. *Practical guide to atmospheric dispersion modeling*. Trinity Consultants, Inc., Dallas, TX.
3. Agency for Toxic Substances and Disease Registry. 1998. Appendix A: Air exposure pathway. In: *Draft public health assessment: Kelly Air Force Base, San Antonio, TX*. Atlanta, GA: ATSDR.
4. Star J, Estes J. 1990. *Geographic information systems: An introduction*. Englewood Cliffs, NJ: Prentice-Hall, Inc. 18.
5. Environmental Systems Research Institute. 1997. Groundbreaking video introducing the world's first GIS now available. *ARCNews* 19(2):33.
6. US Bureau of the Census. 1991. *Census of population and housing, 1990: Summary tape file 1A and summary tape file 3*. Washington, DC: US Bureau of the Census.
7. Agency for Toxic Substances and Disease Registry. 1996. *Air impact assessment during site investigation*. Health assessment training module. Atlanta, GA, June 17–19. Atlanta, GA: ATSDR.
8. Chou, Y. 1997. Surface analysis. In: *Exploring spatial analysis in geographic information systems*. Santa Fe, NM: OnWord Press.
9. Agency for Toxic Substances and Disease Registry. 1993. *ATSDR cancer policy framework*. January. Atlanta, GA: ATSDR.